

G. Surya

Bangalore, Karnataka, India

+91 8667234480 — iamsurya195@gmail.com

linkedin.com/in/g-surya-63a01b290 — github.com/Surya8663

Professional Summary

Computer Science and Engineering undergraduate with applied experience in building and deploying Generative AI systems. Proven expertise in agentic workflows using LangGraph, multi-agent architectures, and production-ready RAG pipelines with Qdrant. Seeking AI Engineer or Software Engineer roles to leverage advanced skills in LLM orchestration, backend engineering, and system design to develop autonomous, scalable AI applications.

Technical Skills

Programming Languages: Python, Java, JavaScript, SQL

AI Orchestration & Frameworks: LangChain, LangGraph, Multi-Agent Systems, RAG, Prompt Engineering

Backend & Web Technologies: FastAPI, Flask, HTML, CSS, REST APIs, Pydantic

Databases: Qdrant (Vector), MySQL, PostgreSQL

Cloud & DevOps: Microsoft Azure, Google Cloud Platform (GCP), Docker, Git

Libraries & Tools: Scikit-learn, Transformers, Pandas, Streamlit, Gradio

Education

B.E. in Computer Science and Engineering — T. John Institute of Technology

Gottigere, Bangalore — Expected Graduation: June 2027

Current CGPA: 8.7/10.0

Relevant Coursework: Data Structures & Algorithms, Software Engineering, Database Management Systems, Computer Networks, Artificial Intelligence

Achievements: Finalist (Top 4 out of 300 teams) in Inter College-Level Hackathon.

Pre-University (XII) — Vijaya Bifurcated PU College

Bangalore, Karnataka — Aug 2021 – Mar 2023

Karnataka Pre-University Board (PCMCs), Grade: 90%

AI Engineering Projects

Agentic RAG Assistant — Python, LangGraph, Qdrant, FastAPI

Jan 2025 – Present

- Built a multi-agent chatbot using LangGraph that routes queries to specialized RAG and web search agents, improving response relevance scores to 0.84.
- Integrated Qdrant vector DB with HNSW indexing for semantic search, reducing average query latency to under 2 seconds.
- **Key Learnings:** State management in cyclic agent workflows, optimizing vector search performance, and designing evaluation metrics for conversational AI.

Aura AI Agent: Profile Semantic Search — Python, FastAPI, Qdrant, RAG

Dec 2024 – Feb 2025

- Developed a RAG system for semantic search across 13,000+ professional profiles, achieving sub-second query response time.
- Implemented data validation and ETL pipelines, ensuring 100% metadata integrity before vectorization.
- **Key Learnings:** Building production-grade RAG pipelines, handling large-scale data ingestion, and creating efficient no-auth API architectures.

AI GitHub Repository Analyzer — Python, FastAPI, LangChain, GitHub API

Nov 2024 – Jan 2025

- Created an automated tool using LLMs to evaluate code quality and skill demonstration in GitHub repos, providing structured feedback in 60 seconds.
- Designed a critique-refine loop to improve analysis depth, reducing shallow feedback by 40%.
- **Key Learnings:** LLM reasoning control for code analysis, structured output parsing with Pydantic, and building developer tools with real-time APIs.

SecurePulse: Behavioral Biometrics — Python, Flask, Scikit-learn

Oct 2024 – Nov 2024

- Built a real-time anomaly detection system using Isolation Forest on user interaction patterns, achieving 95% accuracy in identifying suspicious sessions.
- Developed a Flask backend with a dynamic dashboard for threat visualization and alerting.
- **Key Learnings:** Applying classical ML to security domains, real-time data processing, and building hybrid AI systems that combine statistical models with web applications.

Professional Experience

Product Gen AI Developer intern — HiDevs

Bangalore, Karnataka (Hybrid) — Nov 2025 – Present

- Built Aura AI Agent, a RAG-powered semantic search system over 13,000+ professional profiles with a FastAPI backend and Qdrant vector DB, enabling sub-second retrieval for natural language queries.
- Led the development of ‘Aura AI Agent’ (RAG-based semantic search system) and Github Repository Analyser from concept to deployment in an agile team setting.
- Contributed to architectural decisions and code reviews for multi-agent AI systems, focusing on scalability and performance.

Certifications & Training

Oracle Cloud Infrastructure 2025 Certified AI Foundations Associate – Oracle — Aug 2024

Microsoft Certified: Azure AI Fundamentals – Microsoft — Jun 2024

Foundations of Prompt Engineering – AWS Training & Certification — Aug 2024

Introduction to Prompt Engineering for Generative AI – LinkedIn — Feb 2025

Azure AI Fundamentals Training Program – In-Biot Private Limited — Mar 2024 – Jun 2024

Additional Information

Languages: English (Professional), Kannada (Native), Hindi (Working), Tamil (Working)

Leadership: IIIC Student Coordinator; Volunteer for Quality Circle Forum of India (QCFI) and college placement drives.